

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/103028 A2

- (51) International Patent Classification⁷: C12Q
- (21) International Application Number: PCT/IB02/04189
- (22) International Filing Date: 30 May 2002 (30.05.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/293,999 30 May 2001 (30.05.2001) US
60/330,457 22 October 2001 (22.10.2001) US
60/357,144 19 February 2002 (19.02.2002) US
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant: BIOMEDICAL CENTER [RU/RU]; 7, Pu-
dozhskaja St., 197110 St. Petersburg (RU).
- (72) Inventors: BARANOVA, A. V.; Moscow (RU).
YANKOVSKY, N. K.; Moscow (RU). KOZLOV, A.
P.; St. Petersburg (RU). LOBASHEV, A. V.; Moscow
(RU). KRUKOVSKAYA, L. L.; Moscow (RU).
- Published:
— without international search report and to be republished
upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: IN SILICO SCREENING FOR PHENOTYPE-ASSOCIATED EXPRESSED SEQUENCES

WO 02/103028 A2

(57) Abstract: The present invention provides methods for determining whether a nucleic acid sequence is a marker for a phenotype or cell type of interest which comprises providing a database of expressed sequence tag sequences (EST's) from the species; placing said EST's in groups termed clusters based on homology of EST's within each cluster; determining for each cluster the total number of EST's within said cluster; ordering said clusters sequentially based on the number of EST's in each cluster; dividing said ordered clusters into subranges based on the number of EST's per cluster; determining for each cluster subrange obtained from step (e) the number EST's within said cluster which are expressed in said predetermined cell type of interest; calculating according to a normal distribution the number of clusters in each subrange expected to contain a predetermined threshold percentage of EST's expressed in said cell type of interest, wherein said threshold percentage is a percentage from about 10% to about 100%; determining the number of clusters in each subrange observed to contain said predetermined threshold percentage of EST's expressed in said predetermined cell type; and identifying subranges having an observed number of clusters that meet said predetermined threshold percentage greater than the number of clusters expected to meet said predetermined threshold percentage for the subrange according to normal distribution; wherein if the percentage of EST's expressed in said cell type of interest in a cluster identified is equal to or greater than said predetermined threshold percentage, the cluster contains a nucleic acid that is a marker for the cell type of interest.

Hs.217766	POM105	Ovary carcinoma		SEQ. ID NO: 336	SEQ. ID NO: 337
Hs.217882	POM106	glioma, colon carcinoma, kidney tumors, prostate tumors, lung carcinoma, hypernephroma, head and neck carcinoma, duodenal carcinoma, melanoma, pancreatic carcinoma, uterus tumors		SEQ. ID NO: 338	SEQ. ID NO: 339
Hs.220529	CEACAM5 Carcinoembryonic antigen-related cell adhesion molecule 5	Pancreas carcinoma, colon carcinoma, stomach carcinoma, head and neck carcinoma, lung carcinoma leiomyoma, breast carcinoma	KNOWN TUMOR MARKER	SEQ. ID NO: 340	SEQ. ID NO: 341
Hs.222056	POM107 Homo sapiens cDNA FLJ11572 fig, clone HEMBA1003373	Stomach carcinoma, head and neck carcinoma, breast carcinoma		SEQ. ID NO: 342	SEQ. ID NO: 343
Hs.225083	POM108	Melanoma, ovary tumors, colon carcinoma, parathyroid tumor, kidney tumors, head and neck carcinoma		SEQ. ID NO: 344	SEQ. ID NO: 345
Hs.227098	GCMB Glial cells missing homolog b (Drosophila)	parathyroid tumor		SEQ. ID NO: 346	SEQ. ID NO: 347
Hs.239107	POM109	Lymphoma, germ cell tumors, head and neck carcinoma		SEQ. ID NO: 348	SEQ. ID NO: 349
Hs.239891	GPR35 G protein-coupled receptor 35	B-cell chronic lymphocytic leukemia, colon carcinoma, pancreas and carcinoma	SURFACE	SEQ. ID NO: 350	SEQ. ID NO: 351
HS.241381	CRSP7 Cofactor required for Spl transcriptional activation, subunit 7 (70kd)	Pancreatic carcinoma, duodenal carcinoma, ovary carcinoma, melanoma, osteosarcoma, glioma, leiomyosarcoma, germ cell tumors		SEQ. ID NO: 352	SEQ. ID NO: 353
Hs.241407	SERPINB13 Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 13	ORAL carcinoma, cervical carcinoma, head and neck carcinoma		SEQ. ID NO: 354	SEQ. ID NO: 355
Hs.243920	POM110	Pancreas carcinoma		SEQ. ID NO: 356	SEQ. ID NO: 3
Hs.244378	SLC2A6 Solute carrier family 2 (facilitated glucose transporter), member 6	Hypernephroma, pancreatic carcinoma, glioma, lung carcinoma, neuroblastoma, renal cell carcinoma, adrenal gland tumors		SEQ. ID NO: 358	SEQ. ID NO: 3
Hs.246781	POM111	parathyroid tumor, lung carcinoma tumors, germ cell tumors, hepatocellular carcinoma, stomach carcinoma, breast carcinoma		SEQ. ID NO: 360	SEQ. ID NO: 3